

I. Overview

Fact banding (herein referred to as “banding”) is a method of organizing the appearance of facts (*e.g.*, transaction amount, population census) in reports according to predefined criteria. Banding may be used to organize rows and columns based on either the number of rows or columns, or the values of the row and column headers. Banding has several applications, such as, *e.g.*, organizing income ranges, age ranges and transaction amounts.

In this white paper, four banding methods are presented and analyzed from a performance and flexibility perspective. In the first method, banding is performed using multi-pass SQL. The second method includes performing banding using single-pass SQL with one or more case statements. In the third method, banding is performed using a function that generates an output based on a predefined algorithm. According to the fourth method, herein referred to as “dynamic banding”, banding is performed using a banding lookup table.

II. Background

The four banding methods discussed herein were used to generate reports identical to that displayed in Table 2 using data acquired from the United States Census Bureau. Population data was compiled into approximately four million rows of a census fact table (FACT_CENSUS) displayed in Table 1.

Table 1: Census fact table (FACT_CENSUS).

CENSUS_YEAR	COUNTY_ID	ETHNICITY_ID	GENDER	AGE_GROUP_ID	POPULATION
2000	01001	1	M	1	1186
2000	01001	1	F	1	1200
2000	01001	2	M	1	290
2000	01001	2	F	1	289

FACT_CENSUS included population values for each county included in the census data for the years 2000-2005 (2001-2005 were estimated based on the year 2000 data). For the performance analysis (see section VII), the census data was reinserted into FACT_CENSUS to achieve a table with a desirable number of rows. Where indicated, an index was created on the Population column. While the methods discussed below are presented for population data, the methods can be applied to any dataset in which fact banding is to be applied.

This white paper presents methods for organizing data into a desirable distribution of data ranges. The banding method should ideally be capable of generating the report displayed in Table 2, which is compiled using data obtained from the FACT_CENSUS table.

Table 2: Population counts for predetermined population ranges.

Population Range	Population Occurrence
0-99	3142247
100-199	192782
200-299	119060
300-399	85200
400-499	65409
=500	466018

Banding methods should ideally be capable of grouping data from tables comprising millions of rows while offering the flexibility of having different data range values for each row of a particular data set. As an example, preferred banding methods should be able to generate Table 2 comprising data disseminated into the following data ranges: 0-99, 100-199, etc.